

Data mining

Série N° 2

Exercice 1 : Arbre de décision

Pour mieux appréhender l'induction des arbres de décision, nous allons reprendre un exemple décrit dans l'ouvrage de Quinlan (1993). Il s'agit de prédire le comportement de sportifs (Jouer ; variable à prédire) en fonction de données météo (Ensoleillement, Température, Humidité, Vent ; variables prédictives).

Numéro	Ensoleillement	Température (°F)	Humidité (%)	Vent	Jouer
1	soleil	75	70	oui	oui
2	soleil	80	90	oui	non
3	soleil	85	85	non	non
4	soleil	72	95	non	non
5	soleil	69	70	non	oui
6	couvert	72	90	oui	oui
7	couvert	83	78	non	oui
8	couvert	64	65	oui	oui
9	couvert	81	75	non	oui
10	pluie	71	80	oui	non
11	pluie	65	70	oui	non
12	pluie	75	80	non	oui
13	pluie	68	80	non	oui
14	pluie	70	96	non	oui

L'algorithme d'apprentissage cherche à produire des groupes d'individus le plus homogène possible du point de vue de la variable à prédire à partir des variables de météo. Le partitionnement est décrit à l'aide d'un arbre de décision.

1. Traiter les attributs à valeur continue.
2. Procédez à l'extraction des règles de décisions pour prédire la variable class « jouer ».

Exercice 2 : Arbre de décision

Soit le tableau d'enregistrements suivant :

Fièvre	Douleur	Toux	Maladie
oui	Abdomen	non	Appendicite
non	Abdomen	oui	Appendicite
oui	gorge	non	rhume
oui	gorge	oui	rhume

non	gorge	oui	mal de gorge
oui	non	non	aucune
oui	non	oui	rhume
non	non	oui	refroidissement
non	non	non	aucune

1. Procédez à l'extraction des règles de décisions pour prédire la variable class « maladie »

